Multiple linear regression

Grading the professor

Many college courses conclude by giving students the opportunity to evaluate the course and the instructor anonymously. However, the use of these student evaluations as an indicator of course quality and teaching effectiveness is often criticized because these measures may reflect the influence of non-teaching related characteristics, such as the physical appearance of the instructor. The article titled, "Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity" (Hamermesh and Parker, 2005) found that instructors who are viewed to be better looking receive higher instructional ratings. (Daniel S. Hamermesh, Amy Parker, Beauty in the classroom: instructors pulchritude and putative pedagogical productivity, *Economics of Education Review*, Volume 24, Issue 4, August 2005, Pages 369-376, ISSN 0272-7757, 10.1016/j.econedurev.2004.07.013. http://www.sciencedirect.com/science/article/pii/S0272775704001165.)

In this lab we will analyze the data from this study in order to learn what goes into a positive professor evaluation.

The data

The data were gathered from end of semester student evaluations for a large sample of professors from the University of Texas at Austin. In addition, six students rated the professors' physical appearance. (This is aslightly modified version of the original data set that was released as part of the replication data for *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Gelman and Hill, 2007).) The result is a data frame where each row contains a different course and columns represent variables about the courses and professors.

```
setwd("C:/Users/Christophe/Documents/GitHub/IS606_Fall_2015/Chapter_8/Lab8")
load("more/evals.RData")
```

variable	description
`score`	average professor evaluation score: (1) very unsatisfactory - (5) excellent.
`rank`	rank of professor: teaching, tenure track, tenured.
`ethnicity`	ethnicity of professor: not minority, minority.
`gender`	gender of professor: female, male.
`language`	language of school where professor received education: english or non-english.
`age`	age of professor.
`cls_perc_eval`	percent of students in class who completed evaluation.
`cls_did_eval`	number of students in class who completed evaluation.
`cls_students`	total number of students in class.
`cls_level`	class level: lower, upper.
`cls_profs`	number of professors teaching sections in course in sample: single, multiple.
`cls_credits`	number of credits of class: one credit (lab, PE, etc.), multi credit.
`bty_f1lower`	beauty rating of professor from lower level female: (1) lowest - (10) highest.
`bty_f1upper`	beauty rating of professor from upper level female: (1) lowest - (10) highest.
`bty_f2upper`	beauty rating of professor from second upper level female: (1) lowest - (10) highest.
`bty_m1lower`	beauty rating of professor from lower level male: (1) lowest - (10) highest.
`bty_m1upper`	beauty rating of professor from upper level male: (1) lowest - (10) highest.
`bty_m2upper`	beauty rating of professor from second upper level male: (1) lowest - (10) highest.
`bty_avg`	average beauty rating of professor.
`pic_outfit`	outfit of professor in picture: not formal, formal.
`pic_color`	color of professor's picture: color, black & white.

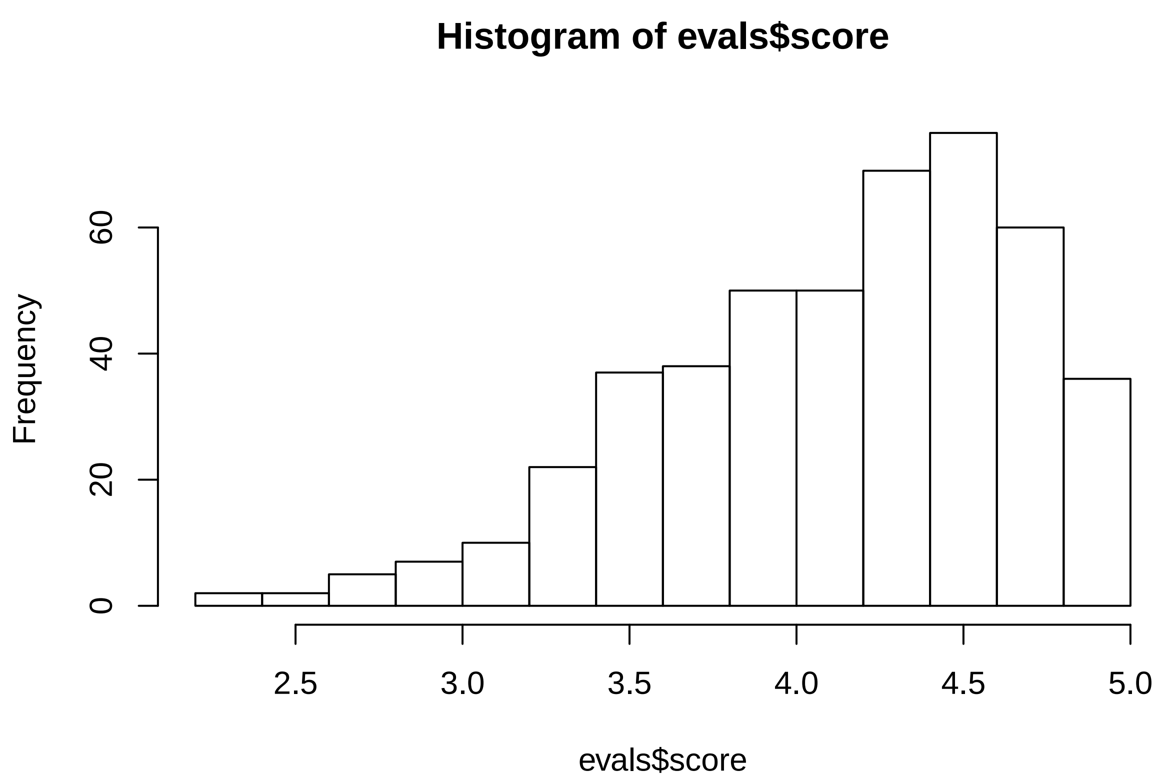
Exploring the data

1. Is this an observational study or an experiment? The original research question posed in the paper is whether beauty leads directly to the differences in course evaluations. Given the study design, is it possible to answer this question as it is phrased? If not, rephrase the question.

   This is an observational study as there was no experiment controlling for other factors. Given that the study was observational it is not possible to infer causality with an observational study. It would be more appropriate to rephrase the question as "Is beauty correlated with differences in course evaluations."

2. Describe the distribution of `score`. Is the distribution skewed? What does that tell you about how students rate courses? Is this what you expected to see? Why, or why not?

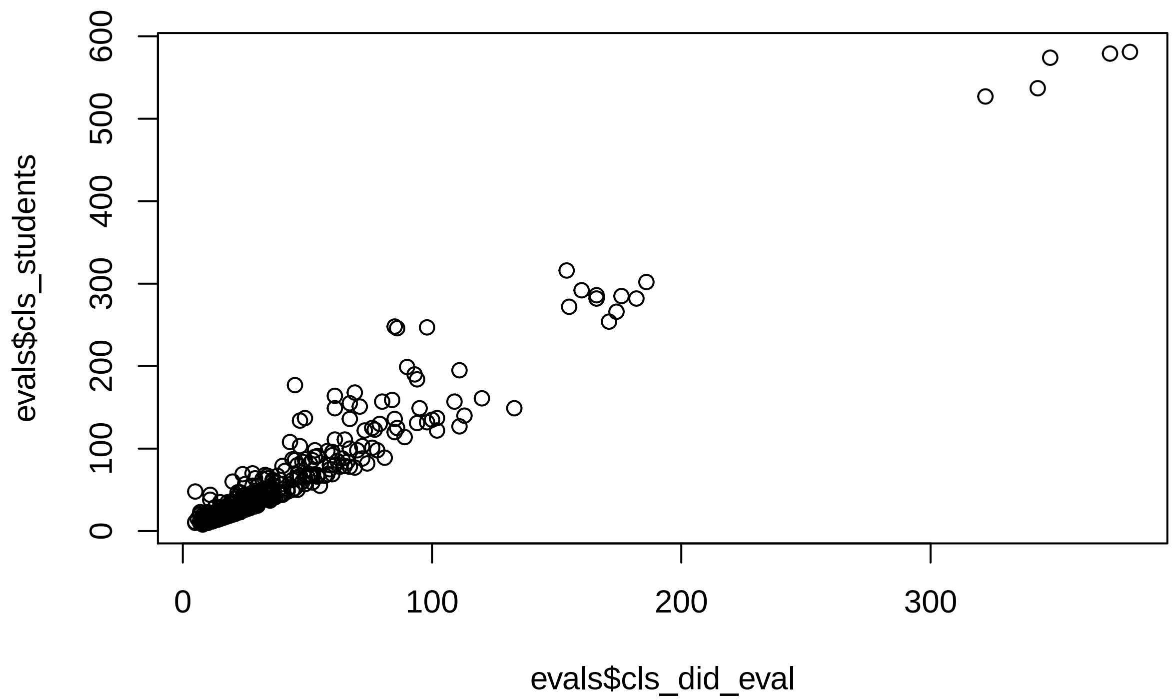
```r
hist(evals$score)
```



   The distribution is unimodal and left skewed. It shows that students typically rate courses fairly high. I would expect this because of natural behavior where people tend to rate things higher. People tend to avoid giving very low scores.

3. Excluding `score`, select two other variables and describe their relationship using an appropriate visualization (scatterplot, side-by-side boxplots, or mosaic plot).

```r
plot(evals$cls_students ~ evals$cls_did_eval)
```

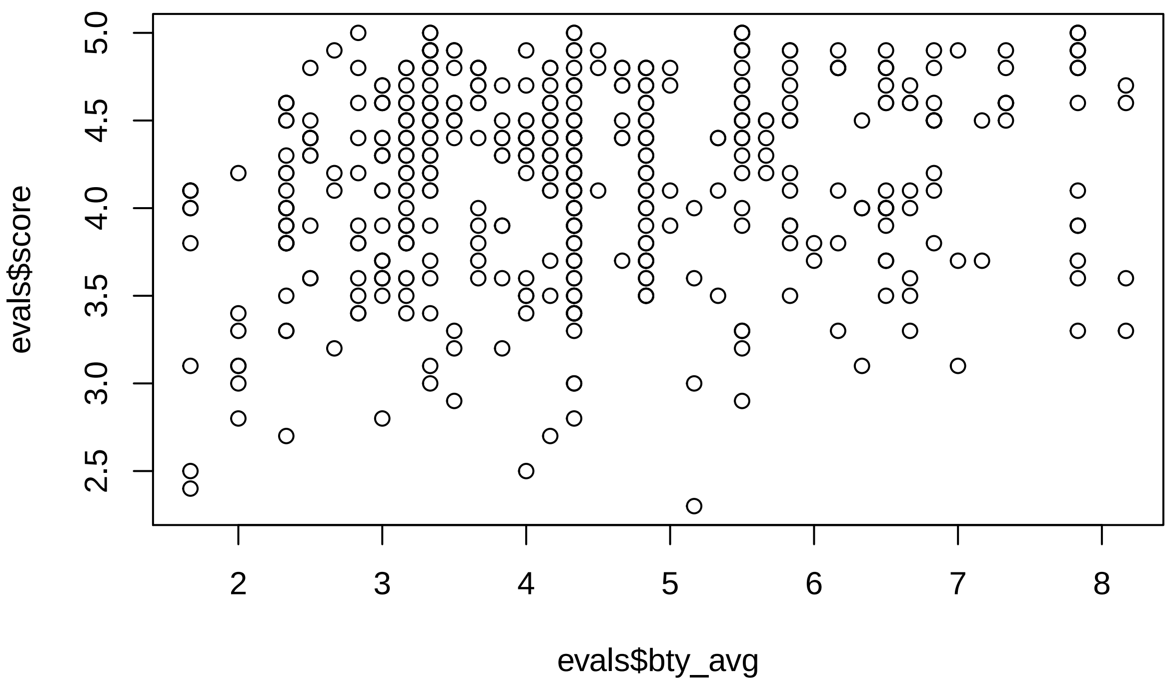


There is a positive correlation between the number of students in the class and the number of students that completed the evaluation. Since we would expect more students completing evaluations for classes with more students this relationship is expected.

Simple linear regression

The fundamental phenomenon suggested by the study is that better looking teachers are evaluated more favorably. Let's create a scatterplot to see if this appears to be the case:

```r
plot(evals$score ~ evals$bty_avg)
```

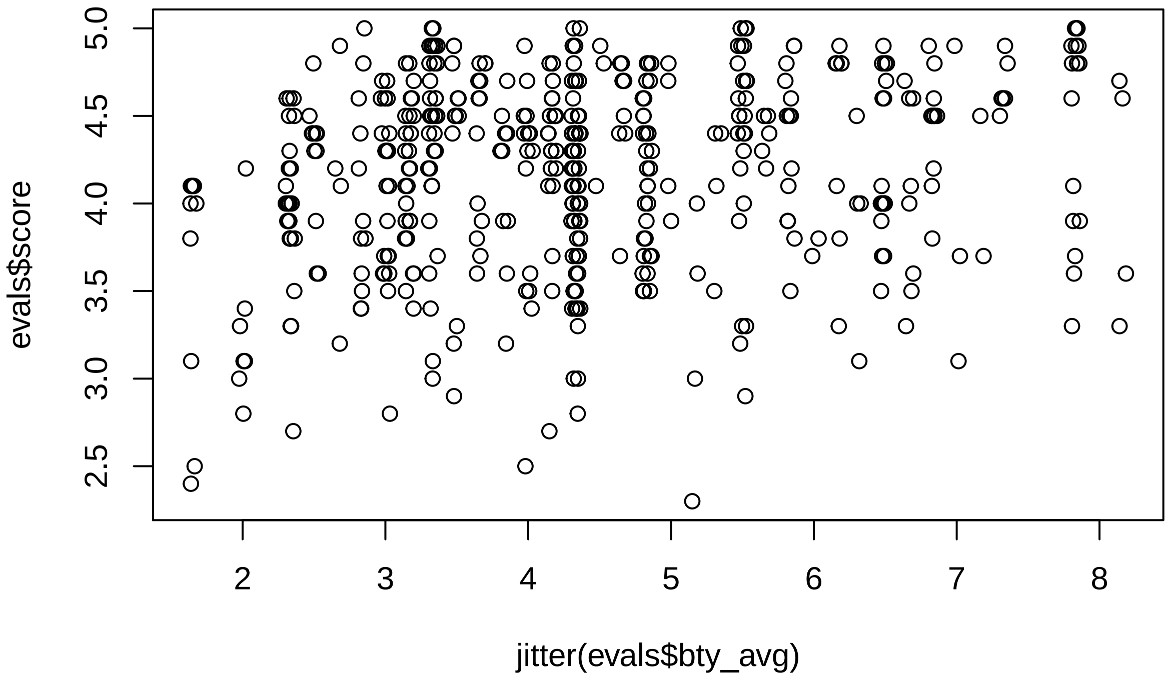


Before we draw conclusions about the trend, compare the number of observations in the data frame with the approximate number of points on the scatterplot. Is anything awry?

Yes, there are far more data points than what is being represented in the above plot.

4. Replot the scatterplot, but this time use the function `jitter()` on the y- or the x-coordinate. (Use `?jitter` to learn more.) What was misleading about the initial scatterplot?

```
plot(evals$score ~ jitter(evals$bty_avg))
```



The previous plot was misleading because it unrepresented the actual observation.

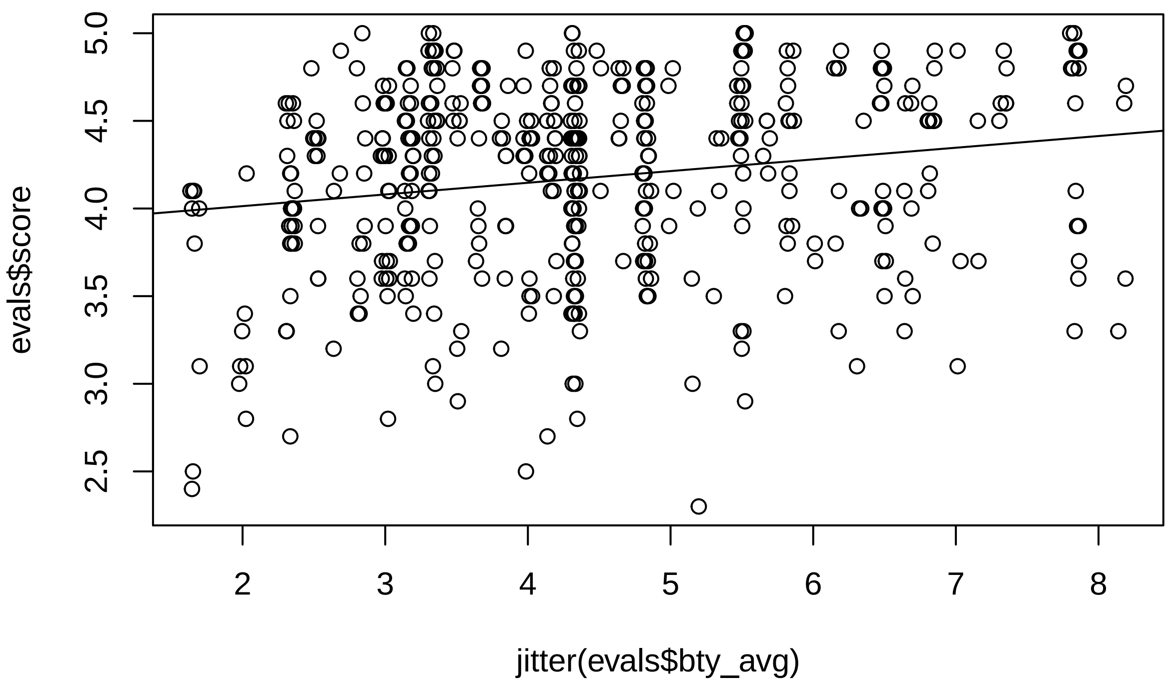
5. Let's see if the apparent trend in the plot is something more than natural variation. Fit a linear model called `m_bty` to predict average professor score by average beauty rating and add the line to your plot using `abline(m_bty)`. Write out the equation for the linear model and interpret the slope. Is average beauty score a statistically significant predictor? Does it appear to be a practically significant predictor?

```
m_bty <- lm(evals$score ~ evals$bty_avg)
summary(m_bty)
```

```
##
## Call:
## lm(formula = evals$score ~ evals$bty_avg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9246 -0.3690  0.1420  0.3977  0.9309
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.88034    0.07614   50.96  < 2e-16 ***
## evals$bty_avg   0.06664    0.01629    4.09 5.08e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.5348 on 461 degrees of freedom
## Multiple R-squared:  0.03502,    Adjusted R-squared:  0.03293
## F-statistic: 16.73 on 1 and 461 DF,  p-value: 5.083e-05
```

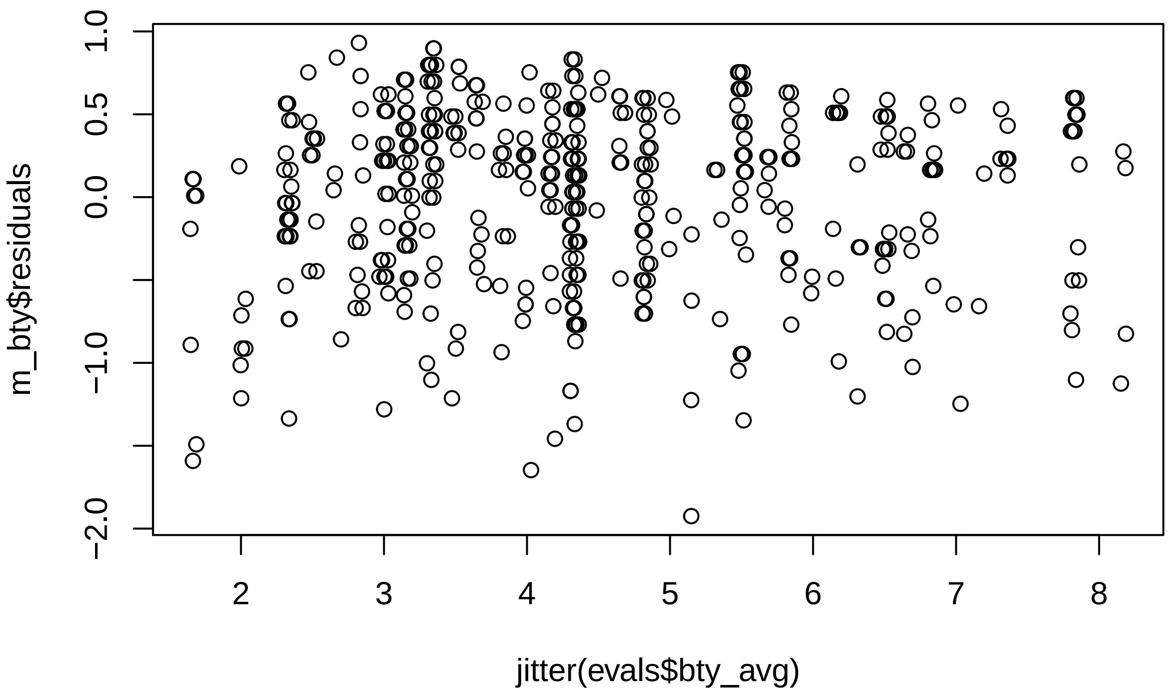
```r
plot(evals$score ~ jitter(evals$bty_avg))
abline(m_bty)
```



The model is $\hat{score} = 3.88 + 0.0666 * bty\ avg$

6. Use residual plots to evaluate whether the conditions of least squares regression are reasonable. Provide plots and comments for each one (see the Simple Regression Lab for a reminder of how to make these).

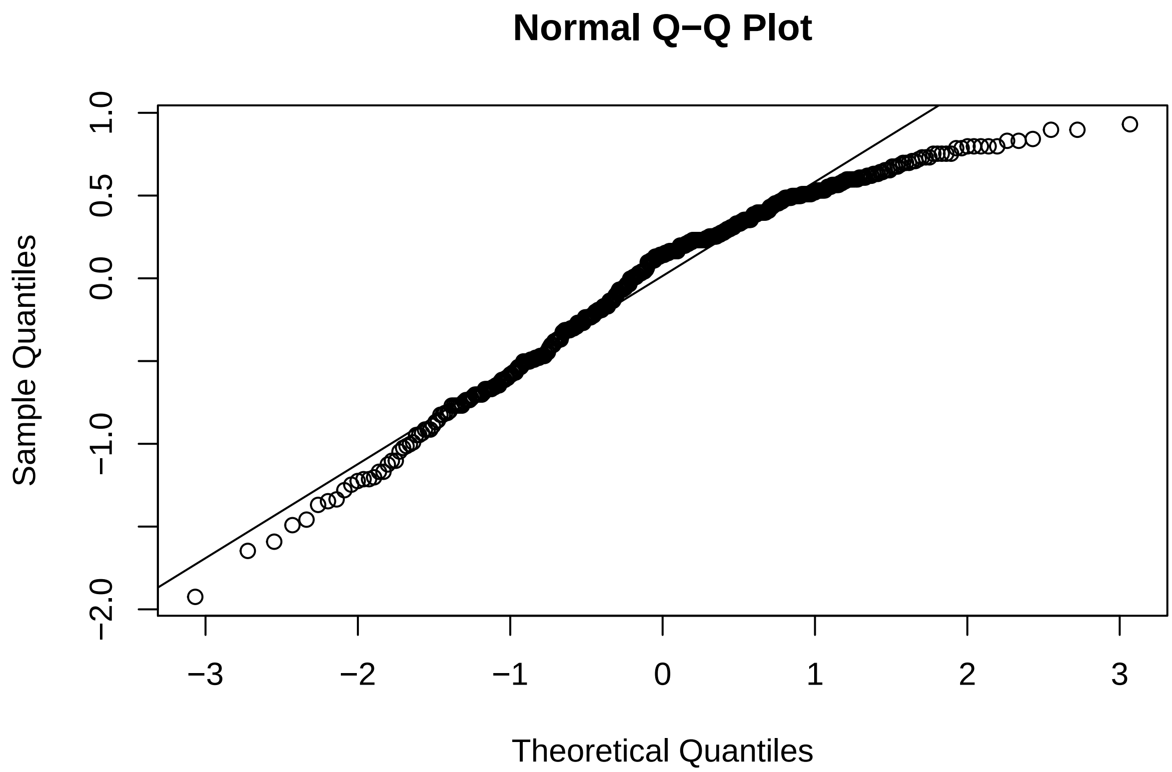
```r
plot(m_bty$residuals ~ jitter(evals$bty_avg))
```




```
sprintf("There does not appear to be any linear relationship to the residuals so the appear near normal
```

[1] "There does not appear to be any linear relationship to the residuals so the appear near normal"

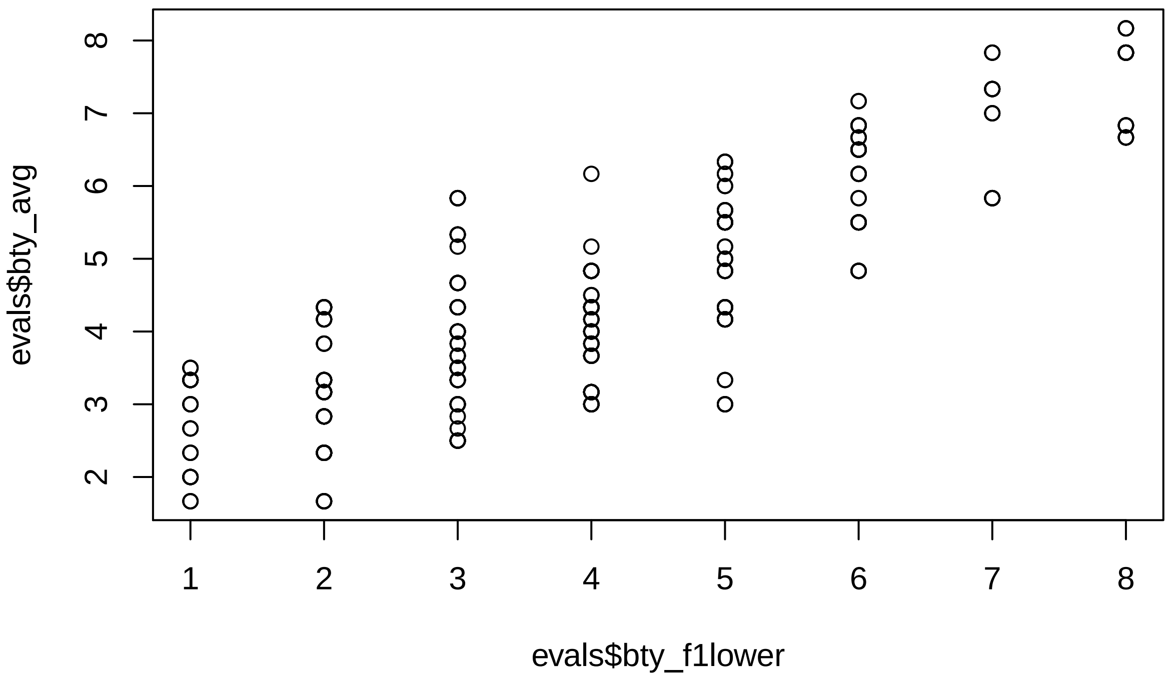
```
qqnorm(m_bty$residuals)
qqline(m_bty$residuals)
```



Multiple linear regression

The data set contains several variables on the beauty score of the professor: individual ratings from each of the six students who were asked to score the physical appearance of the professors and the average of these six scores. Let's take a look at the relationship between one of these scores and the average beauty score.

```
plot(evals$bty_avg ~ evals$bty_f1lower)
```

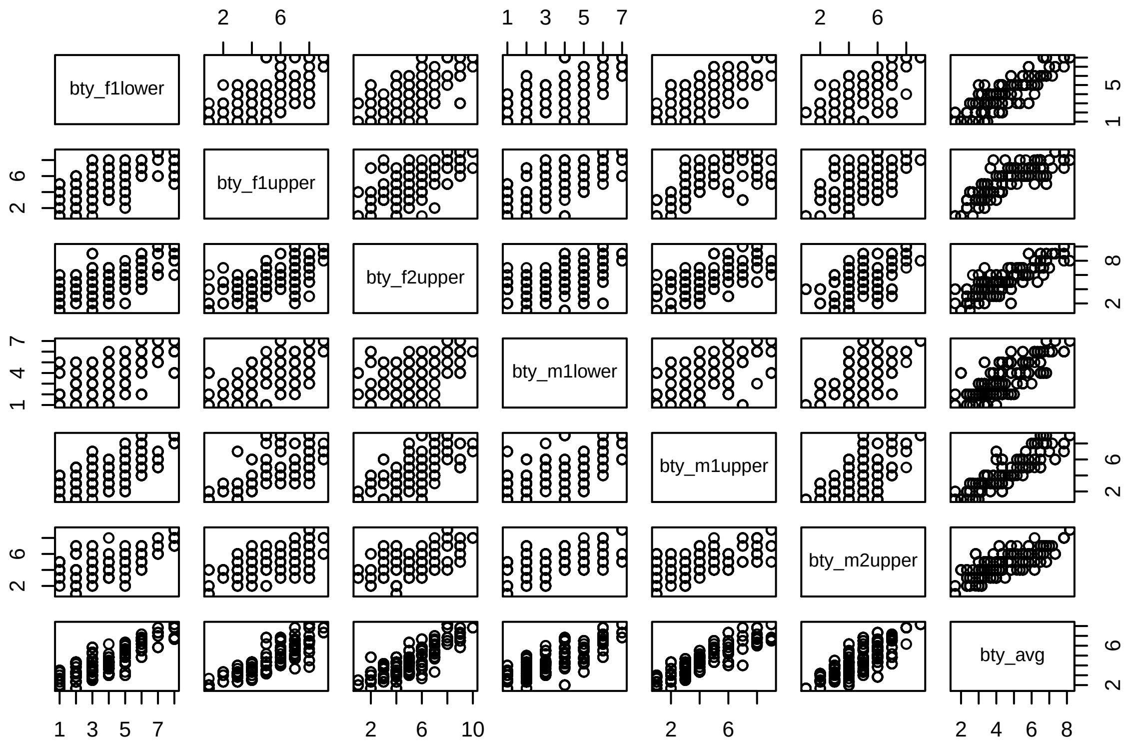



```r
cor(evals$bty_avg, evals$bty_f1lower)
```

```
## [1] 0.8439112
```

As expected the relationship is quite strong - after all, the average score is calculated using the individual scores. We can actually take a look at the relationships between all beauty variables (columns 13 through 19) using the following command:

```r
plot(evals[,13:19])
```



These variables are collinear (correlated), and adding more than one of these variables to the model would not add much value to the model. In this application and with these highly-correlated predictors, it is reasonable to use the average beauty score as the single representative of these variables.

In order to see if beauty is still a significant predictor of professor score after we've accounted for the gender of the professor, we can add the gender term into the model.

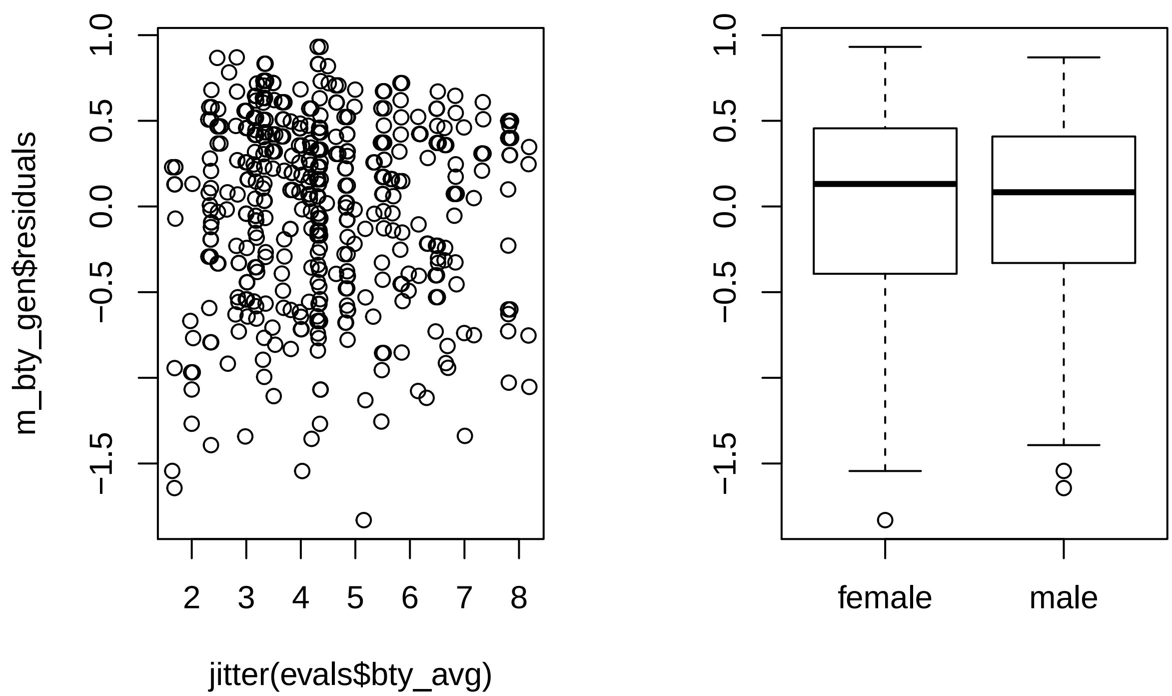
```
m_bty_gen <- lm(score ~ bty_avg + gender, data = evals)
summary(m_bty_gen)
```

```
##
## Call:
## lm(formula = score ~ bty_avg + gender, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8305 -0.3625  0.1055  0.4213  0.9314
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.74734    0.08466  44.266  < 2e-16 ***
## bty_avg      0.07416    0.01625   4.563 6.48e-06 ***
## gendermale   0.17239    0.05022   3.433 0.000652 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.5287 on 460 degrees of freedom
## Multiple R-squared:  0.05912,    Adjusted R-squared:  0.05503
## F-statistic: 14.45 on 2 and 460 DF,  p-value: 8.177e-07
```

7. P-values and parameter estimates should only be trusted if the conditions for the regression are reasonable. Verify that the conditions for this model are reasonable using diagnostic plots.

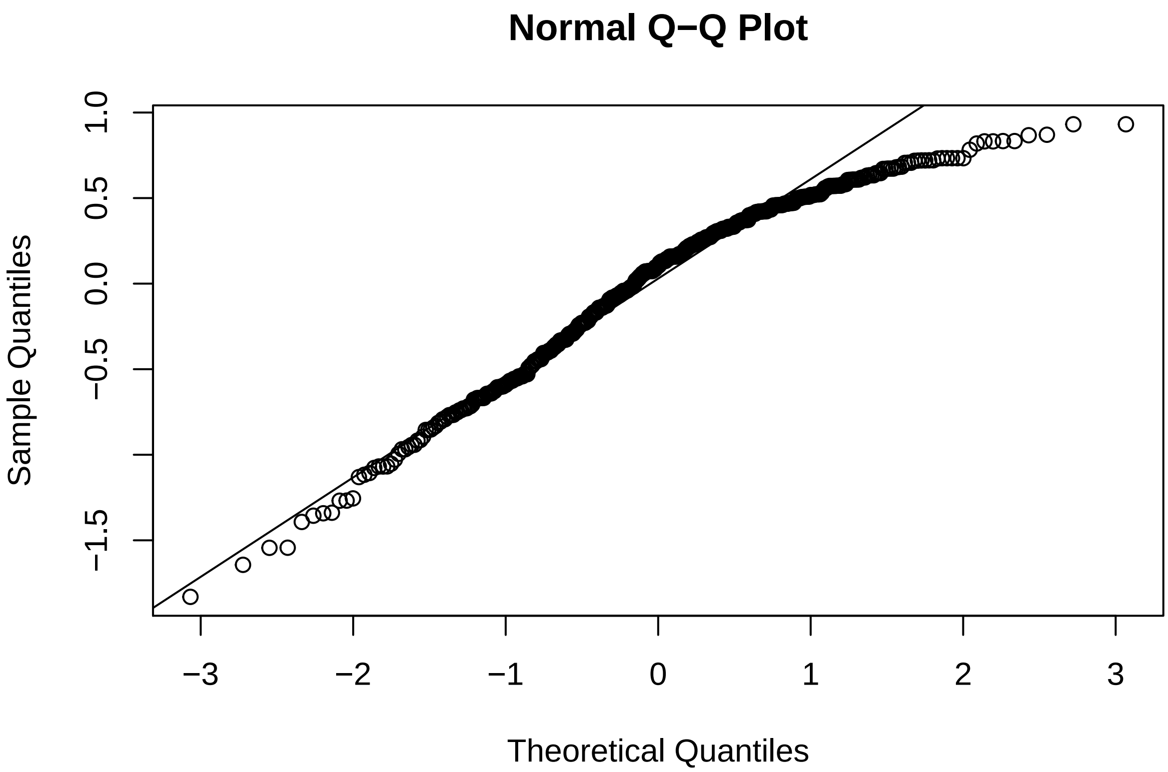
```r
par(mfrow = c(1, 2))
plot(jitter(evals$bty_avg), m_bty_gen$residuals)
plot(evals$gender, m_bty_gen$residuals)
```




```r
sprintf("There does not appear to be any linear relationship to the residuals so they appear near normal
```

[1] "There does not appear to be any linear relationship to the residuals so they appear near normal"

```r
par(mfrow = c(1, 1))
qqnorm(m_bty_gen$residuals)
qqline(m_bty_gen$residuals)
```



8. Is `bty_avg` still a significant predictor of `score`? Has the addition of `gender` to the model changed the parameter estimate for `bty_avg`?

   Yes, because the p value is so low `bty_avg` still a significant predictor of `score`. The addition of `gender` has changed the parameter estimate for `bty_avg` which is expected.

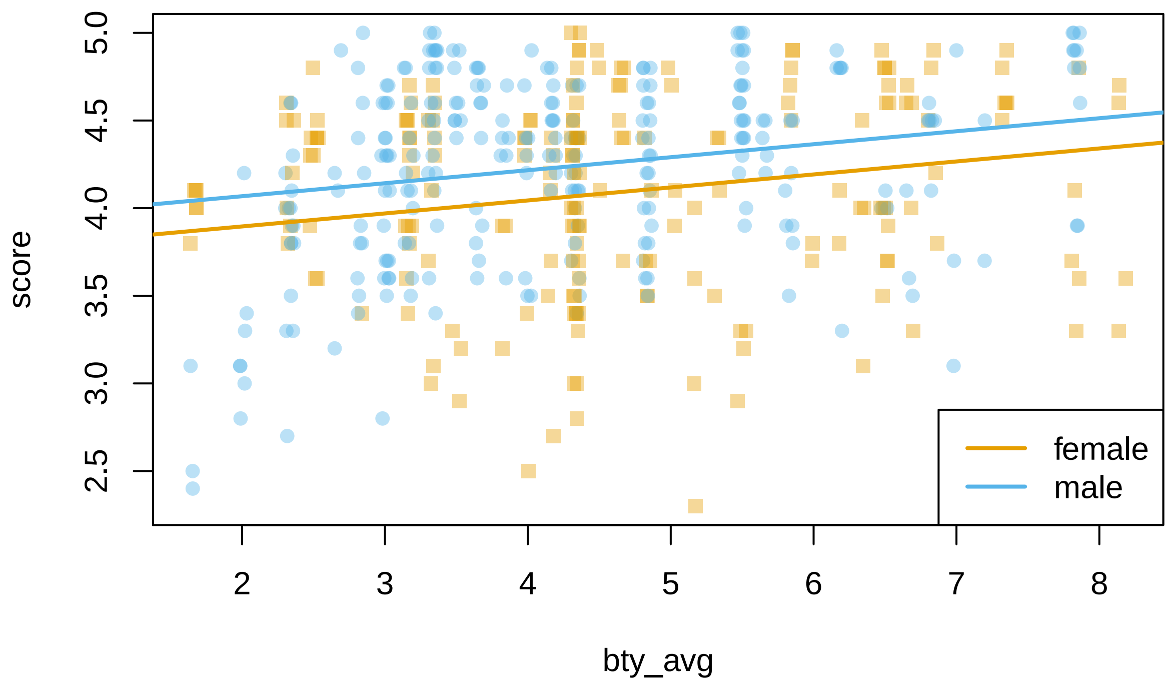
Note that the estimate for `gender` is now called `gendermale`. You'll see this name change whenever you introduce a categorical variable. The reason is that R recodes `gender` from having the values of `female` and `male` to being an indicator variable called `gendermale` that takes a value of 0 for females and a value of 1 for males. (Such variables are often referred to as "dummy" variables.)

As a result, for females, the parameter estimate is multiplied by zero, leaving the intercept and slope form familiar from simple regression.

$$\begin{aligned}\widehat{score} &= \hat{\beta}_0 + \hat{\beta}_1 \times bty\_avg + \hat{\beta}_2 \times (0) \\ &= \hat{\beta}_0 + \hat{\beta}_1 \times bty\_avg\end{aligned}$$

We can plot this line and the line corresponding to males with the following custom function.

```
multiLines(m_bty_gen)
```



9. What is the equation of the line corresponding to males? (*Hint:* For males, the parameter estimate is multiplied by 1.) For two professors who received the same beauty rating, which gender tends to have the higher course evaluation score?

$\hat{score}_{males} = 3.64239 + .07416 * bty\ avg$

According to the model, professors that are male tend of have slightly higher ratings.

The decision to call the indicator variable `gendermale` instead of`genderfemale` has no deeper meaning. R simply codes the category that comes first alphabetically as a 0. (You can change the reference level of a categorical variable, which is the level that is coded as a 0, using the`relevel` function. Use `?relevel` to learn more.)

10. Create a new model called `m_bty_rank` with `gender` removed and `rank` added in. How does R appear to handle categorical variables that have more than two levels? Note that the rank variable has three levels: `teaching`, `tenure track`, `tenured`.

```
m_bty_rank <- lm(score ~ bty_avg + rank, data = evals)
summary(m_bty_rank)
```

```
##
## Call:
## lm(formula = score ~ bty_avg + rank, data = evals)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8713 -0.3642  0.1489  0.4103  0.9525
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.98155    0.09078  43.860  < 2e-16 ***
## bty_avg           0.06783    0.01655   4.098 4.92e-05 ***
## ranktenure track -0.16070    0.07395  -2.173   0.0303 *
## ranktenured      -0.12623    0.06266  -2.014   0.0445 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5328 on 459 degrees of freedom
## Multiple R-squared:  0.04652,    Adjusted R-squared:  0.04029
## F-statistic: 7.465 on 3 and 459 DF,  p-value: 6.88e-05
```

The interpretation of the coefficients in multiple regression is slightly different from that of simple regression. The estimate for `bty_avg` reflects how much higher a group of professors is expected to score if they have a beauty rating that is one point higher *while holding all other variables constant*. In this case, that translates into considering only professors of the same rank with `bty_avg` scores that are one point apart.

The search for the best model

We will start with a full model that predicts professor score based on rank, ethnicity, gender, language of the university where they got their degree, age, proportion of students that filled out evaluations, class size, course level, number of professors, number of credits, average beauty rating, outfit, and picture color.

11. Which variable would you expect to have the highest p-value in this model? Why? *Hint:* Think about which variable would you expect to not have any association with the professor score.

    I would expect `cls_credits` to have the highest p-value because I would assume it would not be able predict the beauty score in any meaningful way.

Let's run the model…

```r
m_full <- lm(score ~ rank + ethnicity + gender + language + age + cls_perc_eval +
    cls_students + cls_level + cls_profs + cls_credits + bty_avg + pic_outfit +
    pic_color, data = evals)
summary(m_full)
```

```
##
## Call:
## lm(formula = score ~ rank + ethnicity + gender + language + age +
##     cls_perc_eval + cls_students + cls_level + cls_profs + cls_credits +
##     bty_avg + pic_outfit + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.77397 -0.32432  0.09067  0.35183  0.95036
##
```

```
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            4.0952141  0.2905277  14.096  < 2e-16 ***
## ranktenure track      -0.1475932  0.0820671  -1.798  0.07278 .  
## ranktenured           -0.0973378  0.0663296  -1.467  0.14295    
## ethnicitynot minority  0.1234929  0.0786273   1.571  0.11698    
## gendermale             0.2109481  0.0518230   4.071 5.54e-05 ***
## languagenon-english   -0.2298112  0.1113754  -2.063  0.03965 *  
## age                   -0.0090072  0.0031359  -2.872  0.00427 ** 
## cls_perc_eval          0.0053272  0.0015393   3.461  0.00059 ***
## cls_students           0.0004546  0.0003774   1.205  0.22896    
## cls_levelupper         0.0605140  0.0575617   1.051  0.29369    
## cls_profssingle       -0.0146619  0.0519885  -0.282  0.77806    
## cls_creditsone credit  0.5020432  0.1159388   4.330 1.84e-05 ***
## bty_avg                0.0400333  0.0175064   2.287  0.02267 *  
## pic_outfitnot formal  -0.1126817  0.0738800  -1.525  0.12792    
## pic_colorcolor        -0.2172630  0.0715021  -3.039  0.00252 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.498 on 448 degrees of freedom
## Multiple R-squared:  0.1871, Adjusted R-squared:  0.1617 
## F-statistic: 7.366 on 14 and 448 DF,  p-value: 6.552e-14
```

12. Check your suspicions from the previous exercise. Include the model output in your response.

    It appear to be wrong as the p value is very low for `cls_credits` which seems unusual. However, the greatest p-value at `0.77806` which can be found for `cls_profs` at the single value. This variable is the the number of professors teaching a course.

13. Interpret the coefficient associated with the ethnicity variable.

    The ethnicity variable indicates that a `not minority` will have a increase likelihood of class score by `.1234` that a minority will not.

14. Drop the variable with the highest p-value and re-fit the model. Did the coefficients and significance of the other explanatory variables change? (One of the things that makes multiple regression interesting is that coefficient estimates depend on the other variables that are included in the model.) If not, what does this say about whether or not the dropped variable was collinear with the other explanatory variables?

    The below tables show the change in values when we remove the highest p-value variable. Some variables begin to explain more of the variance in score and some now explain less.

```r
library(stargazer)
stargazer(m_full, title = 'Regression Results Full Model', header = FALSE)
```

```r
m_less_one <- lm(score ~ rank + ethnicity + gender + language + age + cls_perc_eval
             + cls_students + cls_level + cls_profs +  bty_avg
             + pic_outfit + pic_color, data = evals)

stargazer(m_less_one, title = 'Regression Results Model Less One with Highest P value', header = FALSE)
```

Table 2: Regression Results Full Model

	Dependent variable:
	score
ranktenure track	-0.148^{*}
	(0.082)
ranktenured	-0.097
	(0.066)
ethnicitynot minority	0.123
	(0.079)
gendermale	0.211^{***}
	(0.052)
languagenon-english	-0.230^{**}
	(0.111)
age	-0.009^{***}
	(0.003)
cls_perc_eval	0.005^{***}
	(0.002)
cls_students	0.0005
	(0.0004)
cls_levelupper	0.061
	(0.058)
cls_profssingle	-0.015
	(0.052)
cls_creditsone credit	0.502^{***}
	(0.116)
bty_avg	0.040^{**}
	(0.018)
pic_outfitnot formal	-0.113
	(0.074)
pic_colorcolor	-0.217^{***}
	(0.072)
Constant	4.095^{***}
	(0.291)
Observations	463
R^2	0.187
Adjusted R^2	0.162
Residual Std. Error	0.498 (df = 448)
F Statistic	7.366^{***} (df = 14; 448)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3: Regression Results Model Less One with Highest P value

	Dependent variable:
	score
ranktenure track	−0.196**
	(0.083)
ranktenured	−0.181***
	(0.065)
ethnicitynot minority	0.043
	(0.078)
gendermale	0.237***
	(0.052)
languagenon-english	−0.259**
	(0.113)
age	−0.009***
	(0.003)
cls_perc_eval	0.006***
	(0.002)
cls_students	0.0003
	(0.0004)
cls_levelupper	−0.007
	(0.057)
cls_profssingle	−0.043
	(0.053)
bty_avg	0.032*
	(0.018)
pic_outfitnot formal	−0.136*
	(0.075)
pic_colorcolor	−0.209***
	(0.073)
Constant	4.310***
	(0.292)
Observations	463
R^2	0.153
Adjusted R^2	0.129
Residual Std. Error	0.508 (df = 449)
F Statistic	6.243*** (df = 13; 449)
Note:	*p<0.1; **p<0.05; ***p<0.01

15. Using backward-selection and p-value as the selection criterion, determine the best model. You do not need to show all steps in your answer, just the output for the final model. Also, write out the linear model for predicting score based on the final model you settle on.

```
step(m_full, data=evals, direction="backward")
```

Start: AIC=-630.9 score ~ rank + ethnicity + gender + language + age + cls_perc_eval + cls_students + cls_level + cls_profs + cls_credits + bty_avg + pic_outfit + pic_color

```
        Df Sum of Sq    RSS     AIC
```

- cls_profs 1 0.0197 111.11 -632.82
- cls_level 1 0.2740 111.36 -631.76
- cls_students 1 0.3599 111.44 -631.40
- rank 2 0.8930 111.98 -631.19 111.08 -630.90
- pic_outfit 1 0.5768 111.66 -630.50
- ethnicity 1 0.6117 111.70 -630.36
- language 1 1.0557 112.14 -628.52
- bty_avg 1 1.2967 112.38 -627.53
- age 1 2.0456 113.13 -624.45
- pic_color 1 2.2893 113.37 -623.46
- cls_perc_eval 1 2.9698 114.06 -620.69
- gender 1 4.1085 115.19 -616.09
- cls_credits 1 4.6495 115.73 -613.92

Step: AIC=-632.82 score ~ rank + ethnicity + gender + language + age + cls_perc_eval + cls_students + cls_level + cls_credits + bty_avg + pic_outfit + pic_color

```
        Df Sum of Sq    RSS     AIC
```

- cls_level 1 0.2752 111.38 -633.67
- cls_students 1 0.3893 111.49 -633.20
- rank 2 0.8939 112.00 -633.11 111.11 -632.82
- pic_outfit 1 0.5574 111.66 -632.50
- ethnicity 1 0.6728 111.78 -632.02
- language 1 1.0442 112.15 -630.49
- bty_avg 1 1.2872 112.39 -629.49
- age 1 2.0422 113.15 -626.39
- pic_color 1 2.3457 113.45 -625.15
- cls_perc_eval 1 2.9502 114.06 -622.69
- gender 1 4.0895 115.19 -618.08
- cls_credits 1 4.7999 115.90 -615.24

Step: AIC=-633.67 score ~ rank + ethnicity + gender + language + age + cls_perc_eval + cls_students + cls_credits + bty_avg + pic_outfit + pic_color

```
        Df Sum of Sq    RSS     AIC
```

- cls_students 1 0.2459 111.63 -634.65
- rank 2 0.8140 112.19 -634.30 111.38 -633.67
- pic_outfit 1 0.6618 112.04 -632.93

- ethnicity 1 0.8698 112.25 -632.07
- language 1 0.9015 112.28 -631.94
- bty_avg 1 1.3694 112.75 -630.02
- age 1 1.9342 113.31 -627.70
- pic_color 1 2.0777 113.46 -627.12
- cls_perc_eval 1 3.0290 114.41 -623.25
- gender 1 3.8989 115.28 -619.74
- cls_credits 1 4.5296 115.91 -617.22

Step: AIC=-634.65 score ~ rank + ethnicity + gender + language + age + cls_perc_eval + cls_credits + bty_avg + pic_outfit + pic_color

```
            Df Sum of Sq    RSS     AIC
```

- rank 2 0.7892 112.42 -635.39 111.63 -634.65
- ethnicity 1 0.8832 112.51 -633.00
- pic_outfit 1 0.9700 112.60 -632.65
- language 1 1.0338 112.66 -632.38
- bty_avg 1 1.5783 113.20 -630.15
- pic_color 1 1.9477 113.57 -628.64
- age 1 2.1163 113.74 -627.96
- cls_perc_eval 1 2.7922 114.42 -625.21
- gender 1 4.0945 115.72 -619.97
- cls_credits 1 4.5163 116.14 -618.29

Step: AIC=-635.39 score ~ ethnicity + gender + language + age + cls_perc_eval + cls_credits + bty_avg + pic_outfit + pic_color

```
            Df Sum of Sq    RSS     AIC
```

112.42 -635.39 - pic_outfit 1 0.7141 113.13 -634.46 - ethnicity 1 1.1790 113.59 -632.56 - language 1 1.3403 113.75 -631.90 - age 1 1.6847 114.10 -630.50 - pic_color 1 1.7841 114.20 -630.10 - bty_avg 1 1.8553 114.27 -629.81 - cls_perc_eval 1 2.9147 115.33 -625.54 - gender 1 4.0577 116.47 -620.97 - cls_credits 1 6.1208 118.54 -612.84

Call: lm(formula = score ~ ethnicity + gender + language + age + cls_perc_eval + cls_credits + bty_avg + pic_outfit + pic_color, data = evals)

Coefficients: (Intercept) ethnicitynot minority gendermale
3.907030 0.163818 0.202597
languagenon-english age cls_perc_eval
-0.246683 -0.006925 0.004942
cls_creditsone credit bty_avg pic_outfitnot formal
0.517205 0.046732 -0.113939
pic_colorcolor
-0.180870

```
best_model <- lm(formula = score ~ ethnicity + gender + language + age + cls_perc_eval +
    cls_credits + bty_avg + pic_outfit + pic_color, data = evals)
stargazer(best_model, title = 'Regression Results For Best Model', header = FALSE )
```

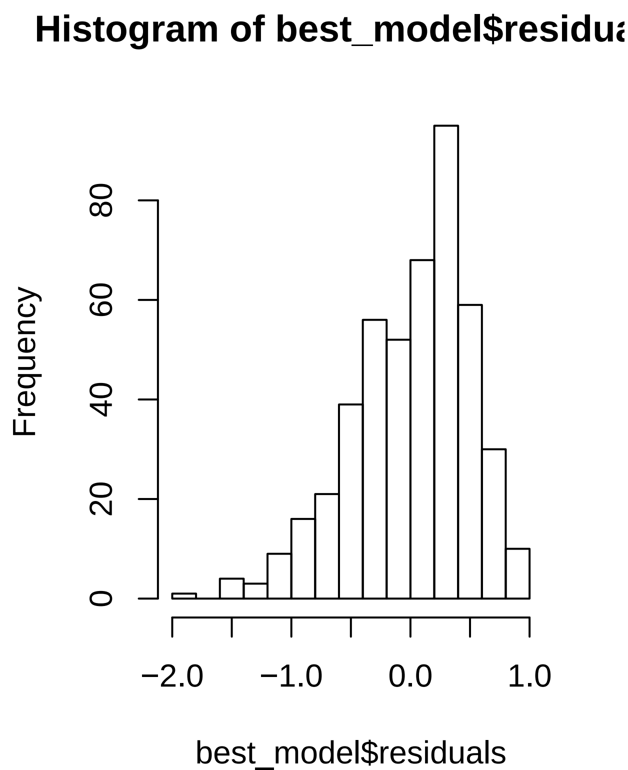
$\hat{score} = 0.164 * ethnicity_{not\ minority} + .203 * gender_{male} - 0.247 * language_{non-english} - 0.007 * age + .005 * cls\ perc\ eval + .517 * cls\ credit_{one\ credit} - .046 * bty\ avg - 0.114 * pic\ outfit_{not\ formal} - .181 * pic\ color_{color} + 3.907$

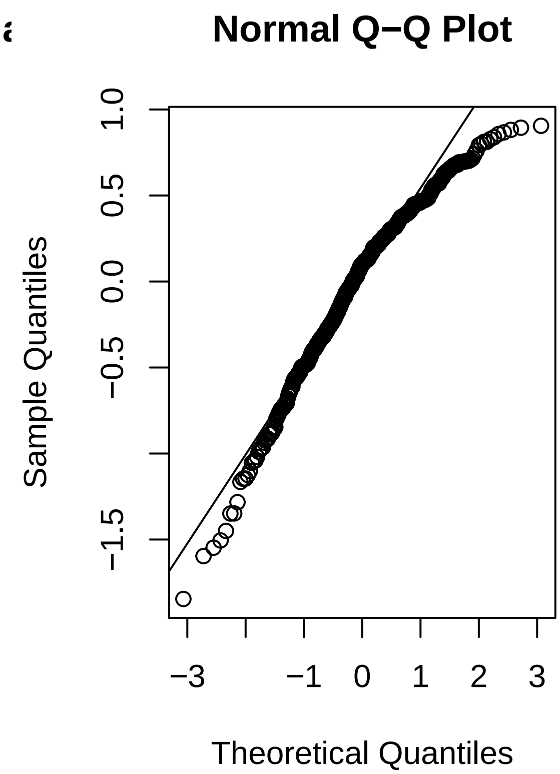
Table 4: Regression Results For Best Model

	Dependent variable:
	score
ethnicitynot minority	0.164**
	(0.075)
gendermale	0.203***
	(0.050)
languagenon-english	$-$0.247**
	(0.106)
age	$-$0.007***
	(0.003)
cls_perc_eval	0.005***
	(0.001)
cls_creditsone credit	0.517***
	(0.104)
bty_avg	0.047***
	(0.017)
pic_outfitnot formal	$-$0.114*
	(0.067)
pic_colorcolor	$-$0.181***
	(0.067)
Constant	3.907***
	(0.245)
Observations	463
R^2	0.177
Adjusted R^2	0.161
Residual Std. Error	0.498 (df = 453)
F Statistic	10.853*** (df = 9; 453)
Note:	*p<0.1; **p<0.05; ***p<0.01

16. Verify that the conditions for this model are reasonable using diagnostic plots.

```
par(mfrow = c(1, 2))
hist(best_model$residuals)
qqnorm(best_model$residuals)
qqline(best_model$residuals)
```





The conditions for this modely are reasonable as the historgram does not show significant skew and the quartiles plot follows the residual line.

17. The original paper describes how these data were gathered by taking a sample of professors from the University of Texas at Austin and including all courses that they have taught. Considering that each row represents a course, could this new information have an impact on any of the conditions of linear regression?

    This could impact the independence of the observations. If they selected many courses from certain professors there may be some unintentional weights added. That is, if one professor is sampled many times while another professor is only represented a few times we may see bias in the results.

18. Based on your final model, describe the characteristics of a professor and course at University of Texas at Austin that would be associated with a high evaluation score.

    A professor with a high estimated score would not be a minority, male, english speaker, younger with a high percentage of students completed the evaluations. Also, they would be teaching a one credit course, with a low beauty average, and a formal picture that was in black and white.

19. Would you be comfortable generalizing your conclusions to apply to professors generally (at any university)? Why or why not?

    I would not be comfortable generalizing my conclusions to professors at any university. The reason being that I am not entirely comfortable with the selection process for evaluation. I am not able to confirm how professors and courses were selected for this study. It seems that many confounding factors could be in play that I am not aware of so I would not be comfortable applying this to any university.